# Morphology in Second Language Acquisition: the Acquisition of English Compounds

**Balajić, Marta**

Sveučilište u Zadru

Odjel za anglistiku
Sveučilišni prijediplomski studij Anglistike

**Marta Balajić**

# Morphology in Second Language Acquisition: the Acquisition of English Compounds

**Završni rad**

Zadar, 2023.

Sveučilište u Zadru

Odjel za anglistiku

Preddiplomski sveučilišni studij engleskog jezika i književnosti (dvopredmetni)

# Morphology in Second Language Acquisition: the Acquisition of English Compounds

Završni rad

Student/ica:

Marta Balajić

Mentor/ica:

Doc. dr. sc. Frane Malenica

Zadar, 2023.

# Izjava o akademskoj čestitosti

Ja, **Marta Balajić,** ovime izjavljujem da je moj **završni** rad pod naslovom **Morphology in Second Language Acquisition: the Acquisition of English Compounds** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 21. rujna 2023.

**Table of contents :**

## 1. Introduction

Language acquisition refers to the process of learning and achieving the knowledge of a language and the ability to understand and employ its principles in order to produce and comprehend it adequately. Morphology, as a study of word forms and their formation, has extensively been analysed in the context of language acquisition. This paper deals with the acquisition of compound words in a second language, compounding being one of the most productive word formation processes in English. The second chapter of this paper explains the basic postulates of morphological theory in language acquisition through a psychological lense and explains how first language might affect the acquisition of new languages. The third chapter then presents a classification of English compound words, their processing, and a comparison of their features in English, Croatian and Spanish. Finally, the fourth chapter is a description and analysis of a research conducted with participants, native speakers of Croatian and Spanish that tests the influence of leisure time activities on the acquisition of English compound words.

## 2. Acquisition of morphology

This chapter deals with the theoretical background of language learning in psycholinguistics from a morphological point of view. The second section offers a comparison between inflection and derivation regarding morphological processes in language learning, and lastly, the third section explains how different languages interfere in a speaker's mind and discusses some issues concerning second language acquisition.

### 2.1. Morphological processing in psycholinguistics

Every speaker of a human language has a mental lexicon. It is individual, it does not have a limited number of entries and it comprises both passive and active vocabulary, the one used in language production. The words that make up the mental lexicon are called lemmas. Lemmas are defined as lexical entries that support the surface manifestation of actual lexemes and they comprise sociolinguistic, syntactic, morphological, semantic, and syntactic information that is crucial to a speaker choosing one lemma over another (Wei, 2002, p. 692). The process of storing words in our minds is incessant; there are always new words added and words which are not being used anymore. Unlike dictionaries, mental lexicons do not have a threshold when it comes to coining new words. A lot more time and linguistic procedures are needed for a word to enter a dictionary. Second of all, the words inside a mental lexicon are connected on a phonological, morphological, and semantic levels, which in a dictionary is not expressed directly and words have only one type of relationship which is the alphabetical order (Booij, 2007, p. 233). Human minds cannot work out a frequency number in which a certain word appears in a language, and neither can a dictionary, but language speakers can make valid assumptions saying if one word is used more frequently than another one. For example, an English speaker could easily say that the word *'nice'* is used far more frequently than *'ubiquitous'*. The frequency at which a word appears in a given language determines the speed at which it is activated in the mind.

The mental lexicon is computed by what is called morphological knowledge. Morphological knowledge refers to the set of inflectional and derivational rules and individual complex words stored in our mind. Both native and non-native speakers of a given language develop their language skills based on this knowledge. When it comes to the question of whether a speaker is aware or unaware of using morphological knowledge, it has been divided into explicit and implicit knowledge. Implicit knowledge is subconscious grammatical knowledge that a speaker uses automatically and the explicit one pertains to the conscious knowledge of grammatical rules that a speaker is able to explain (Friedline, 2011, p. 22).

Another division of the morphological knowledge that concerns the direction in which words are processed is receptive and productive morphological knowledge. Receptive knowledge refers to recognizing the meaning of a word by dividing it in its constituent parts, and when speakers use productive morphological knowledge they create, rather than receive, the meaning of morphologically complex words and use them in a context accordingly (Friedline, 2011, p. 18).

The two main ways in which morphological knowledge is used are computing or production and retrieving words from memory. By retrieving words from memory, we use the mental lexicon in which words are stored as they are memorized. According to the frequency effect, as the frequency of occurrence of a word rises so does the speed at which it is processed. On the other hand, computation implies using morphological rules to compute the meaning of a word or produce a new one by dividing it in its constituent parts – morphemes and making use of the knowledge and memory of inflectional and derivational rules and affixes. Thus, processing different words in a language largely depends on a speaker's morphological ability. According to Libben (2012), morphological ability is the ability of a speaker to use prefixes and suffixes accordingly in a way that allows creating new forms while at the same time maintaining the forms that already exist in the mind. He divides it into four abilities: the ability to repeat, comprehend, and produce multimorphemic words with appropriate semantic and syntactic properties, the ability to understand novel multimorphemic constructions, the ability to produce novel multimorphemic constructions and the ability to employ morphological patterns within the language in order to organize vocabulary knowledge (Libben, 2012, cited in Archibald and Libben, 2019, p. 532-533).

Ullman (2004) introduced the terms of declarative and procedural systems as the two main systems used to employ the mental lexicon and grammatical rules. The lexical or declarative model involves the retrieving words from memory or the mental lexicon. The procedural or grammatical model, however, corresponds with making use of the mental grammar in coining new words and word forms. He argues that these two systems interact in multiple ways, such as that in some cases they can be used for acquiring the same type of knowledge (Ullman, 2004, p. 247).

### 2.2. Inflection vs. derivation in morphological processing

Inflection and derivation are the two main morphological processes in the creation of new words and word forms. In inflectional processes, one lexeme changes across grammatical categories such as gender, number, or tense. The result of inflection are different word forms

of the same lexeme which belong to the same word class. In derivation, however, new lexemes are created by means of derivational processes such as affixation or compounding, the result of which are words that belong to different word classes.

It has been argued that in inflectional morphology there are three stages of morphological acquisition, for both first and second language learners, which create a U-curve (Booij, 2007, p. 237). Firstly, it is supposed that both regular and irregular forms of a word are learnt by heart and used accordingly. In the second stage a rule is recognized by the speaker, but it is not used properly as, for example, irregular forms are assigned inflectional affixes of regular forms. For example, English past tense form of the verb *'write'* will be formed as regular *'writed'* instead of irregular *'wrote'*. Overgeneralization, also known as the final stage, occurs when the speaker has acquired both rules and exceptions. It is believed that there are two different types of processing routes of regular and irregular forms. A number of linguists believe that the acquisition of regular and irregular forms is processed by a dual system model, which means that regular forms are, when needed, computed in the moment of speaking by using morphological rules and grammar and speakers memorize the irregular forms and store them in the mind. However, there have been debates on the influence of the frequency of occurrence of regular forms and whether they are produced on the spot or also memorized because of their high frequency and activation level. Hence, other linguists propose single-route processing as a tenable explanation for the acquisition of inflectional forms. In single-route processing both regular and irregular forms are learnt by heart (Blom, 2019, p. 515).

There have been several research studies on how inflectional forms are processed in L1 and L2 learners, on the order and the difficulty of their acquisition, and inflectional morphology as an individual system in language learning. With respect to derivational morphology, less studies have focused on its processing in a second language, and it has not been observed in the same way that inflectional morphology has. One of the reasons might be that derivational morphological ability in L2 learners is not as easily observed (Archibald & Libben, 2019, p. 532). Another possible reason is the Split Morphology Hypothesis as proposed by Anderson (1982) and Perlmutter (1988), who argue that derivational processes only happen within the lexicon and that inflection governs the computation of words by means of grammatical rules (Anderson, 1982; Perlmutter, 1988, cited in Friedline, 2011, p. 35). In other words, that derived words are exclusively stored and inflected words are computed, which according to some empirical studies in psycholinguistics is not viable and they claim that derivation and inflection could be computed within the same system of rules (Friedline, 2011, p. 36).

The main process by which derived words are processed in language learning is morphological decomposition. It consists of dividing multi-morphemic words into their constituent parts and in that way processing their meaning. For example, removing the prefix or the suffix from a derived word to compute the meaning of the word base. This is not to say that speakers can not process derived words as whole words as well, though. The ability of a speaker to morphologically decompose complex words and in that way understand and process their meaning is referred to as receptive knowledge, as described in chapter 2.1. There are two main theories concerning lexical processing of derived words. Namely, those two are dictionary metaphor and network theories (Baayen, 2014). The dictionary metaphor theory uses the dictionary as an example of how the mental lexicon is organised and how it is used. Network theories, on the other hand, explain and describe lexical processing in terms of order of its activation, which goes from form units to semantic units, and is not based on entry forms such as those in dictionaries. In theories building on the dictionary metaphor, the processing of complex words is significantly influenced by the frequency of words. The higher the frequency of the word, the faster it is activated. A number of research show that L2 acquisition in naturalistic settings shares many similarities with L1 acquisition in the sense that the more a speaker is encountered with a word, the faster it will be processed (Elley & Mangubhai, 1983, p. 54). Thus, it is assumed that the number of hours of interaction with a language highly influences language acquisition and processing. Typically the difference in the amount of hours of exposure to L1 and L2 is immense, so trying to reduce the hour gap could also reduce the difference in competence level (Elley & Mangubhai, 1983, p. 55).

The issue with the frequency effect here is that with the increase of the number of lexical entries the search for them becomes more challenging. Because of this, according to this theory, vocabulary units are grouped into word families based on their constituents. Thus, the search for words in the mental lexicon supposedly has two parts. The first search finding the correct word family to which the word belongs, and the second search that involves looking for the meaning inside the word family itself (Baayen, 2014). The frequency theory then applies to constituents in derived, multi-morphemic words as well. If the first constituent is the one searched first in the mental lexicon, the more frequent it is, the faster it will be processed. It should still be noted that the frequency of the complex word itself does affect constituent frequency in that complex words with high frequency enable higher activation level of their constituent parts.

A problem concerning morphological parsing is that it is supposed that speakers use it subconsciously and automatically. What happens is that in the search for possible word base or constituent parts, complex words can be incorrectly decomposed. For example, in the processing of derived prefixed words, the prefix is sometimes removed from unprefixed words. This results in meaningless word strings such as *cipice* and *ique* after incorrectly removing the prefix from the words 'precipice' and 'unique' (Taft, 1981, cited in Baayen 2014). Because of how misleading obligatory morphological decomposition can be, it is argued that its disadvantages outweigh the advantages (Baayen, 2014).

### 2.3. Language interference and some issues in second language acquisition

As was mentioned in the first section of this chapter, all language speakers have a mental lexicon. The mental lexicon can be monolingual, bilingual, or multilingual, depending on how many languages a person speaks. The way that lexical entries enter and function inside a mental lexicon works the same in first, second language acquisition and in the acquisition of all additional languages a speaker comes to be surrounded with. The question that rises when words from other languages enter the mental lexicon is whether they are all organised inside a single mental lexicon, or whether each language is organised in an individual lexicon. Linguists generally believe that bilinguals possess a single mental lexicon in which, the activation of L1 and L2 lexical items varies depending on the language being spoken at the time (Wei, 2002, p. 693). This is to say that items that belong to different languages are activated in different contexts, and when words from one language are activated, the ones from the other language are automatically deactivated. Because of this it becomes irrelevant whether there is a single or multiple mental lexicons since the vocabulary from different languages functions on the basis of its activation in the moment of speaking. The key issue then appears to be the relationship between the lexical items in these two languages and whether they are in any way connected or activated in synchrony. There are linguists who claim that both L1 and L1 items can, in fact, be activated at the same time. The simultaneous activation, according to word recognition tests conducted by Beauvillain and Grainger (1987) and Grainger and Dijkstra (1992), only happens in the initial lexical access and lemmas from the two languages are not necessarily always equally active. This means that both lexicons are initially activated, and language-specification process happens afterwards when some lemmas from that language may be more activated than their counterparts in the other language (Beauvillain and Grainger, 1987, Grainger and Dijkstra, 1992, cited in Wei, 2002, p. 694).

Although sometimes languages a person speaks have significant differences in their morphological structure, 'the essential architecture and processing mechanisms of monolinguals and bilinguals are the same' (Libben and Goral, 2015, cited in Archibald and Libben, 2019, p. 528). Morphological knowledge works on the same principles in first and all other languages a person acquires. What is debatable here is the extent to which native and non-native speakers of a given language use the declarative and procedural system in language processing. In other words, how much non-native speakers tend to use the declarative and procedural systems when compared to native speakers.

The results of a masked-priming study conducted by Silva and Clahsen (2008) showed that morphology is processed differently by L2 learners than by native speakers. According to what the priming effects indicated, L2 learners relied less on decomposition and were unaware that, for example, numerous adjectives can have the suffix *-ness* added to them to create nouns like *darkness*, *awareness*, and *disease* (Silva and Clahsen, 2008, cited in Friedline, 2011, p. 38). Another experimental research, by Clahsen and Neubauer (2010), showed that L1 speakers of, in this case German, showed that frequency affected only the irregular forms, and L2 speakers showed it for both irregulars and regulars. This showed that L2 processing is more dependent on lexical memory than L1 processing. The results from both research show that, in second language acquisition, drawing back on memory storage is more salient in language processing than morphological decomposition and using the procedural system is (Friedline, 2011, p. 39). Ullman (2005) supports this claim by agreeing that 'L2 acquisition will tend to rely more on the declarative system than L1 acquisition does' (Ullman, 2005, cited in Archibald and Libben, 2019, p. 534).

Despite what was found in this research, it is still possible that L2 speakers of a language will rely more on morphological decomposition in language reception and production because they are familiar with less multimorphemic words than native speakers and they are often in need of coining new word-forms that they possibly have not encountered before. Therefore, when encountered with an unfamiliar multimorphemic word, they will rely on their knowledge of morphological rules and stem forms in their mental lexicon in the comprehension of the word. Under this assumption, even though second language learners may primarily rely on morphological processing, they are at the same time less proficient at it than native speakers. (Archibald and Libben, 2019, p. 536).

In investigating second language acquisition it is of vital importance to consider the morphological structure of the languages that interfere in the speaker's mind and the way

morphological differences in the new language are acquired. In 1960s Noam Chomsky introduced the term Universal Grammar, an important postulate in linguistics that proposes the existence of a predefined mechanism in the human brain that is the basis for the acquisition of all languages. He proposed this as an idea of innate, biological grammatical categories, such as the noun and verb categories, which are crucial for language development in children and adult processing. Universal Grammar encompasses all the grammatical information required to combine e.g., nouns and verbs, into phrases (Nowak et al., 2001, p. 115). Universal Grammar does not, however, explain the acquisition of grammar of particular languages.

A new language a speaker is encountered with may have some morphemes that are absent in their L1 or miss certain morphological features that are present in their L1. It is not an issue for learners to notice these differential features in the morphological structure, but it is a challenge when it comes to acquiring them (Schmidt, 1990, cited in Archibald & Libben, 2011, p. 528). That is, the morphemes that are absent from their L1 should be learnt and, more importantly, tagged appropriately (Archibald & Libben, 2019, p. 527). Language tagging refers to the process of marking and organising lexical entries in the mental lexicon in different categories and contexts. By means of this feature of the mental lexicon, certain language items are deactivated in specific conversations or contexts, as explained at the beginning of this section. For example, if a bilingual is using their L2 in the moment of speaking, the lexical items from their L1 will automatically be deactivated because they are not tagged for the context (Archibald & Libben, 2019, p. 527). Hence, in morpheme acquisition, once learnt, the morphemes of the new language should be tagged correctly for the morphological transfer to be completed.

Archibald and Libben (2019) use the example of how Chinese speakers acquire English tense morphology in explaining the acquisition of new morpheme features of a language. Chinese is a language that does not have tense marking like English, but instead expresses the tense feature by marking the aspect. In English, the verb eat would take the past form 'ate' in expressing the past tense [+past], while in Chinese the same feature would be expressed by adding the word '*le*' which indicates that an action is completed [+perfective] (Archibald & Libben, 2019, p. 528). The two languages use different morphological structures to express the same linguistic meanings. Thus, Chinese learners must not only learn the past tense feature in English but also associate it with the appropriate equivalent in their native language.

A problem that might appear in the investigation of speakers' acquisition of morphology is the uncertainty of whether a speaker commits errors because a grammatical feature was not

acquired properly or because they are unable to demonstrate it in the performance or language production (Archibald & Libben, 2019, p. 524). A speaker's poor performance in subject-verb agreement does not necessarily indicate a lack of knowledge of abstract agreement in their second language. Often, they are aware of the underlying 'rule' in the process even though they struggle to apply it when speaking spontaneously (Friedline, 2011, p. 22). The representational problem has frequently been analysed in the context of the omission of some morphemes in language acquisition due to the observation that 'second language learners often omitted inflectional morphemes in L2 production' (Archibald & Libben, 2019, p. 523). In the example of Chinese learners of English, Hawkins and Chan (1997) claim that if a representational feature, like [±past], is absent from one's L1, it will not be acquired in their L2 and assign to this the problem of the omission of tense morphology (Hawkins and Chan, 1997, cited in Archibald & Libben, 2019, p. 525).

# 3. Compound words

This chapter delves into the theoretical overview of compound words and how they are acquired in language. After a classification of compounds presented in the first section, the paper examines the acquisition of compound words and, lastly, analyses how ambiguous compound words are processed.

## 3.1. Typology and types of compounds

Compounding is a highly productive word formation process in which two or more lexemes are combined to create new ones. The constituent parts of compound words can belong to different word classes, and what determines which word class the compound as a whole belongs to is the head of the compound, which is one of the constituents, or both of them. If they are single-headed, compounds can be right-headed or left-headed, while the majority of English compounds are the former (Fabb, 2017, p. 67). The non-head constituent is most commonly referred to as the modifier. When it comes to the lexical meaning of its constituents and the compound word itself, they can either coincide or not. Considering all of the features of compound words, they have been classified in many ways by linguists in the past and it is still an ongoing debate (Bisetto & Scalise, 2005, p. 319). In this paper a classification proposed by Bisetto & Scalise (2005) is presented.

Bisetto & Scalise (2005) propose a division of compound words, firstly according to the grammatical relation between its constituents into subordinate, coordinate and attributive, which can then all be divided into endocentric and exocentric, based on the presence or absence of a head constituent (Bisetto & Scalise, 2005, p. 326). In subordinate compounds one of the constituents is the modifier of the other one, which is the head. The relation of constituents in subordinate compounds is thus called complement relation. Attributive compounds combine a noun and an adjective or two nouns which are also in a modifier-head relation where the modifier is an attribute of the head and expresses one of its properties. Finally, the constituents in coordinate compounds have a head-head relation in which the heads are virtually connected by the conjunction 'and'. As for the notion of endocentricity and exocentricity, endocentric compounds have a head constituent that carries the grammatical and semantic meaning of the entire compound and is the hyponym of the entire compound. Exocentric compounds, on the other hand, lack a constituent that is a formal head and are therefore not hyponyms of their head (Bisetto & Scalise, 2005, p. 327).

As has already been mentioned, the meaning of the compound word and its constituents can vary. Linguistic terminology of this factor and an important feature of compound words is

semantic transparency, or contrary, semantic opacity. The transparency of a compound word is determined by the degree to which the compound's overall meaning and its constituents' meaning are interconnected. If the interpretation of a compound word's meaning can be determined by examining the lexical meaning of its constituents, such as in *blackboard*, it is defined as fully semantically transparent. If the whole-word meaning of a compound is, however, incompatible with the meaning of its constituents and there is not a clear relationship between them, it is called a semantically opaque compound word, such as *hogwash* (Steinke & Zhai, 2020, pp. 2). Compounds can also be partially opaque, such as *butterfly*, with one constituent having unrelated meaning (butter) and the other linked to the whole-word meaning (fly).

### 3.1.1. Compounding in English, Croatian and Spanish

English, Spanish and Croatian are languages that belong to three different Indo-European language groups. Namely, to Germanic, Romance and Slavic languages. In all three languages compounds as a word form exist, but they use compounding to different extents. Romance and Slavic languages in general resort to compounding much less frequently than Germanic languages (Renner, V. & Fernández-Domínguez, J., 2011, p. 3). English is a language in which compounding is a highly efficient word-formation process. On the other hand, compounding in Spanish is much more restricted, with less than 425 compound words in the database (Sebastián-Gallés, Martí, Carreiras, & Cuetos, 2000, cited in Duñabeitia, Perea, & Carreiras, 2007, p. 1172). Although compounding in Croatian is a common word formation process, affixation is much more productive (Grčević, M., 2015, p. 3000). Since English compounds can be written as one word, hyphenated or as a phrase and Croatian grammar does not regard phrases and hyphenated expressions as compounds, English covers a wider range of structures and therefore has a larger number of compounds overall (Štambuk, A. 1997, p. 374).

While in English and Spanish compounds are divided into endocentric and exocentric, in Croatian linguistic literature this division, although present, is not common since most Croatian compound words are endocentric. Endocentric compounds are generally considered to be a more recent phenomenon in Slavic languages than exocentric compounds (Grčević. M., 2015, p. 3001). One of the most notable differences between Spanish and Croatian and English compounds is that in Croatian and English, endocentric compounds are most frequently right-headed and in Spanish they are mostly left-headed (Piera, 1995, p. 4). Another difference in compounding between these languages is that in English it is recursive and in Croatian and Spanish it is not (Piera, 1995, p. 4). All of the factors mentioned make it apparent that

compounding is a much more productive and flexible process in English than it is in Croatian and Spanish.

### 3.2. Ambiguous novel compounds

It is apparent that when a speaker is encountered with a novel compound, or a compound that is not familiar to them because of its low frequency in language, they will try to analyse its meaning by looking at the meaning of the constituent morphemes, although it does not always guarantee a correct interpretation. It presents a challenge the fact that the constituents of compound words belong to an open class of words, and the possibilities of their combinations are infinite (Libben et al., 2006, p. 7). In affixation, for example, it is much easier to strip off the affixes in morphological parsing, which belong to a closed class, and be left with a string that most probably has a single lexical entry in the mental lexicon (Taft and Forster, 1975, cited in Libben, 2006, p. 7). Since a large portion of compound words cannot be processed only by examining the individual morphemes, lexical storage is essential in their processing and acquisition. The mental lexicon's representation of compound words remains unclear since they can be represented by their constituents or as a whole through a separate entry (Bell & Schäfer, 2016, p. 159). One thing we can be sure of is that semantically opaque compounds should always have an individual lexical representation because if not, the speaker will not be able to process them correctly.

It is difficult to know the amount of effort semantically opaque compounds require from a speaker to be processed. While a lot of them do require more time and the conjoint use of morphological parsing and lexical storage to be processed, speakers are often not even aware of the fact that some compounds they use on a daily basis are even compound words, especially that they are semantically opaque. For example, some common fruit and vegetable compounds like grapefruit, strawberry or pineapple seldom draw attention to their semantic opacity because of how well-known and frequent in language they are (Libben et al., 2006, p. 7). Nonetheless, when it comes to novel semantically opaque compounds, especially ambiguous ones, they take much longer to process. A question that needs to be answered here is what factors affect speakers' choices in the morphological parsing of novel compounds that can be decomposed in two ways. Libben (1994) and Libben et al. (1999) conducted a set of experiments in which they investigated how ambiguous novel compounds like *clamprod*, *seathorn* and *cartrifle*, are parsed. The participants were instructed to read aloud the aforementioned words listed in a compound stimulus that also contained other non-ambiguous compound words. Their dependent variable was the location of the pause the participants would make in speech,

representing their interpretations of morpheme boundaries inside the words. What they found was that the semantic plausibility, rather than the frequency of the first constituent, was what was correlated with the parsing choices. This means that both parsing possibilities were observed, but the one where the constituents seemed to be a better semantic fit was chosen. If both possibilities were, in fact, observed in cases with ambiguous compounds, it would then be logical to assume that they required longer processing time than other compounds from the stimuli. This was additionally proven in an experiment with a lexical decision task where 30 subjects needed a significantly larger amount of time to reject ambiguous novel compounds as real words of English compared to unambiguous ones (Libben et al., 2006, p. 9-10).

### 3.2.1. Processing of compound words

Although quite a narrow field in linguistics, the analysis of compound word processing is fundamental in the investigation of lexical processing and psycholinguistics (Libben et al., 2006, p. 1). What makes them so important in linguistic research of word acquisition is the fact that they are multi-morphemic and consist of individual lexemes, which means they can be studied as whole words or morphologically decomposed and analysed respectively (Libben et al., 2006, p. 2). It is questionable the extent to which morphological computation and lexical storage are used in their analysis because of the nature of their word class and meaning, defined by headedness and semantic transparency.

Headedness significantly impacts the processing of compound words. First of all, the head is the element that indicates what part of speech the word belongs to. For example, the word class of the word '*blackboard*' is a noun because its head is the noun *board* and not the adjective *black*. Secondly, studies have shown that the heads of compound words have a greater activation level than their other constituent (Jarema et al., 1999, cited in Steinke & Zhai, 2020, p. 2). In a masked-priming experiment conducted by Steinke & Zhai (2020), compounds primed with head-related words showed a faster reaction time than those primed with words related to their nonhead constituent (Steinke & Zhai, 2020, p. 7).

Semantic transparency is arguably one of the most important factors in the processing of compound words and has considerably been investigated in linguistic research. As explained in the previous section, semantic transparency refers to how much constituent meaning aids the understanding of the meaning of the whole compound (Bell & Schäfer, 2016, p. 158). The processing of semantically opaque compound words has been of special interest because constituent meaning in this case typically does not contribute to the whole-word meaning, but is arguably still activated in the mind. It is the parallel activation of constituent and whole-word

meaning that makes their analysis unique. This then makes semantically opaque compounds more challenging to process than semantically transparent ones, since the speaker activates constituent meaning which does not benefact the understanding of the whole word, but on the contrary impairs the process of its understanding.

In explaining how whole-word and constituent interpretation possibly cooperate in the analysis of the meaning of semantically opaque compounds, Libben (2006) describes a case study conducted with a patient with impaired judgement due to a brain injury, who was asked to interpret a number of opaque compound words including *butterfly*, *yellowbelly*, and *dumbbell*. The patient's paraphrases of them are presented in the following table.

**Table 1.** Patient's interpretations of opaque compound words (Libben, 2006)

| COMPOUND WORD | PATIENT'S PARAPHRASE |
|---|---|
| butterfly | 'a pretty fly…it's yellow' |
| yellowbelly | 'a yellow stomach…a chicken' |
| dumbbell | 'stupid weights' |

As can be seen in Table 1., the patient interpreted both whole-word and constituent meaning. She identified the meaning of *butterfly* by saying it is a pretty insect but mentioning that it is yellow probably shows access to the literal meaning of the constituent '*butter*'. In *yellowbelly* she first identified the independent meaning of the constituents (a yellow stomach), and then interpreted it as 'a chicken', which is connected to the whole-word meaning 'coward'. Lastly, associating *dumbbell* with 'stupid' and 'weights' also shows access to both constituent and whole-word meaning. Although using both whole-word and constituent meaning impaired her interpretation, it is possible that this was not the problem. If non-impaired English speakers also accessed both meanings in the interpretation of compounds, as an automatic and obligatory process, it would mean that they also automatically choose the right option, which is what the patient in this case study failed to do (Libben, 2006, p. 3-4). This does mean, however, that semantically opaque compounds take longer time to process than semantically transparent ones.

## 4. Acquisition of compound words in a second language: a research

I conducted a research exploring the difference in the understanding of some ambiguous and transparent compound words from English between native speakers of Croatian and Spanish, and how certain external factors affect this knowledge. In the following sections I list the hypotheses and research questions, explain the methodology used and present the results of the research.

### 4.1. Hypotheses and research questions

One of the hypotheses I wanted to test in this research was that speakers of both Croatian and Spanish who spend more hours interacting with English would perform better in the compound word recognition test because of the higher frequency of occurrence of these words in their surrounding which provokes a higher activation level in their mental lexicons. Analogously, those with less interaction hours were expected to perform less well. It was also expected that the proficiency level would affect the performance in that the participants with higher proficiency level would show better results in the word recognition test. Age and gender were not expected to affect the performance in any aspect. I also hypothesized that subjects would show a better performance with semantically transparent compounds than with semantically opaque ones and that words with higher frequency would be recognized more than words that have lower frequency in the corpus.

Research questions I intended to answer with this research are:

1. Do native language and time spent reading in English as an L2 affect speakers' knowledge of compound words?
2. Does the difference in native language influence participants' accuracy in compound word test?
3. Does semantic transparency of the compounds and their frequency have an influence on their recognition as English words by the participants?

In other words, whether leisure time and the subjects' L1 affect L2 performance and what the difference in this performance is like between Croatian and Spanish native speakers in recognizing English compounds.

### 4.2. Methodology

A research was conducted to examine the effect of interacting with the English language in leisure time activities on the knowledge and understanding of English compound words by Croatian and Spanish EFL learners. The participants were 24 Croatian and 21 Spanish EFL

learners in the age group of 20 to 30 year olds that have been learning English for at least 10 years. The instrument used was a survey consisting of three parts. The first part concerned leisure time activities and the amount of time spent reading content in English in different media. This part was followed by 20 multiple-choice questions about compound words. Half of the compounds were semantically transparent and the other half semantically opaque ones, with different frequencies. I extracted the word frequency values from COCA (Corpus of Contemporary American English). The last part of the survey included a sample SAT evidence-based reading and writing task to measure the approximate proficiency level of each participant. The survey contained 50 questions in total. I calculated the final test score of the SAT sample test by interpreting them according to the SAT scoring process and got final results which vary from 200 to 800 points. A potential problem was the motivation of the participants to solve the tasks and the fact that they probably did not double-check their answers, as they would normally do with tests that are to be graded. It is also important to note that there was no time limit for either part and that the measure in this research was an offline measure – the accuracy of answers in the tests. Therefore, for the results of this research to be fully reliable it would have been best to use an inline measure (reaction time) in the compound word recognition test and to examine the proficiency level in the SAT by solving the complete evidence-based test with a time limit. This would take longer and would make the data-collection process more challenging, but the results would be more genuine. Another aspect that could not be analysed because of the missing time limit is the processing time for the compound words which would show which compounds take longer to process than the others. What the survey did show however, is how successful the participants were in identifying the correct forms of 20 English compounds presented in the task.

I administered the survey online via Google Forms and extracted the results to Microsoft Excel. I used a sample evidence-based reading and writing test from Khan Academy which has practice tests available for everyone. I did not take the full test, but only one text from each section and questions referring to them, respectively. After looking at the size of the dataset and using measures of central tendency to check whether it is normally distributed, I removed the rogue values by marking them ‘NA’. All of the inferential statistics tests were conducted in JASP.

### 4.3. Results and analysis

There were 45 respondents to the survey, three of which did not consent to the use of their data for the purposes of the research and did not fill out the rest of the form. 46

observations were collected from each participant, including the questions from the sample test. After grading them and calculating the scores, there were 25 individual observations from the first and the second part, which makes a total of 1050 observations. Two datapoints were considered rogue values because of an extreme number of hours per week listed in the part with leisure time activities. There were no outliers in the dataset. A mean score of 428 was obtained in the sample SAT evidence-based reading and writing task with a standard deviation of 121. The dataset I got was normally distributed.

To begin with, I created a linear regression model to assess the effect of the number of hours interacting with English in leisure time on the results in compound word recognition test. Compound word test results were entered as the dependent variable and the number of hours as the independent variable. The model showed a statistically significant result and the number of hours proved to affect the results in the test ($F_{(1, 38)}$= 13.886, p<.001). It accounted for a significant proportion of variance in the test results ($R^2$=.268). The intercept of the model was 12.877 and the slope was 0.058.

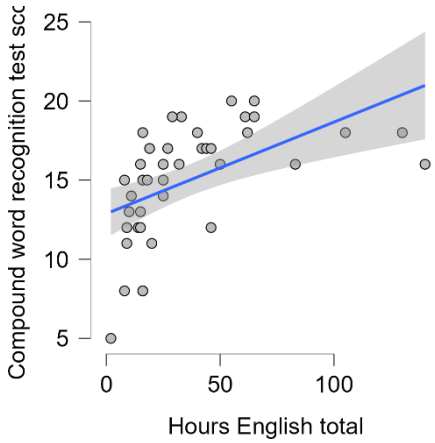


**Figure 1.** Scatter Plots: Hours English total - Compound word recognition test score

As can be seen, the participants who spend more time interacting with and using English in their leisure time were more successful in the compound words test in the research, which confirms my first hypothesis.

Another linear regression model was created to analyse a potential relationship between the proficiency level (recorded in the SAT sample test) and the compound word test scores. The dependent variable was the proficiency level, and the results of compound word test was the independent one. The model was shown to be statistically significant

($F_{(1, 38)}$=33.522, p<.001). These results show that the proficiency level has a positive effect on participants' successfulness in compound word test. There is a significant proportion of variance in the test ($R_2$=.469). The intercept of the model was 6.271 and the slope was 0.020.

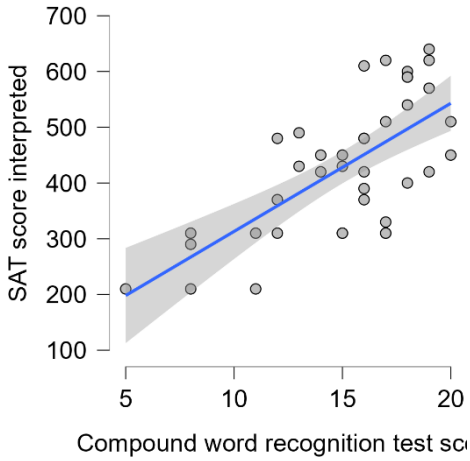


**Figure 2.** Scatter Plots: Compound word recognition test score – SAT score interpreted

The next step was conducting a chi-squared test to examine the correlation between semantic transparency and the accuracy of answers in the compound words test. A statistically significant result was established between the two variables ($\chi 2$= 9.460, df = 1, p<.05, N = 640). The association between the variables was low and they were shown to have a negative relationship since the Phi-coefficient is -0.122. A conclusion can be drawn from this that, as expected, semantically transparent compound words were recognized more by the participants.

**Table 2.** Contingency tables: Semantic opacity and accuracy of answers

| | | Semantic opacity | | |
|---|---|---|---|---|
| Accuracy | | opaque | transparent | Total |
| correct | Count | 235.000 | 267.000 | 502.000 |
| | Expected count | 251.000 | 251.000 | 502.000 |
| incorrect | Count | 85.000 | 53.000 | 138.000 |
| | Expected count | 69.000 | 69.000 | 138.000 |
| Total | Count | 320.000 | 320.000 | 640.000 |
| | Expected count | 320.000 | 320.000 | 640.000 |

A linear regression model was used in the analysis of the correlation between compound word frequency (obtained from a corpus source) and the accuracy of answers in the compound words test. Although the result was not shown to be statistically significant ($F_{(1,17)}$=2.471, p>.05), it can be seen in Table 5. that there is a trend in the correlation between these two variables and that the results show a higher accuracy rate with compound words that have higher frequency in the corpus. Since the result here was not statistically significant, it did not answer one of my research questions.

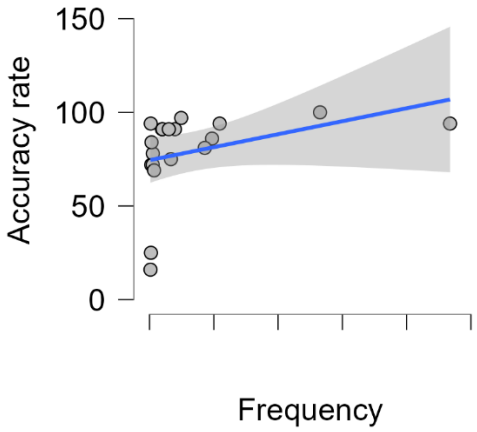


**Figure 3.** Scatter Plots: Frequency - Accuracy rate

In the analysis of the effect of native language on speakers' knowledge of compound words in this research, I conducted a t-test, which showed a statistically significant result (t=3.159, df=40, p<.05) that indicates that native speakers of Croatian performed better than the Spanish, as shown in Figure 4. The strength of association was shown to be small (η2=0.2). The mean that Croatian participants obtained in the compound word recognition test was 16.6 and the Spanish mean was 13.45. This could indicate that native language was the factor that affected the speakers' knowledge of compound words in the test. However, it should be taken into account that Spanish participants, in comparison to Croatians, showed a lower overall number of hours that included interacting with the English language in their leisure time activities (Figure 5.). It is also important to mention that Croatians obtained better SAT sample test results ($M_{Croatian}$=453, $M_{Spanish}$=382). Thus, the results of the t-test do not and can not indicate that the nature of the native language was the determining factor in the compound word test performance.

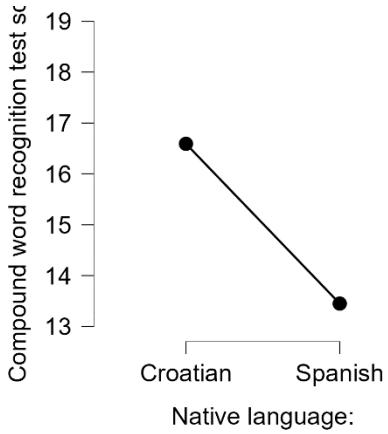


**Figure 4.** Descriptives plots: Compound word test score and Native language

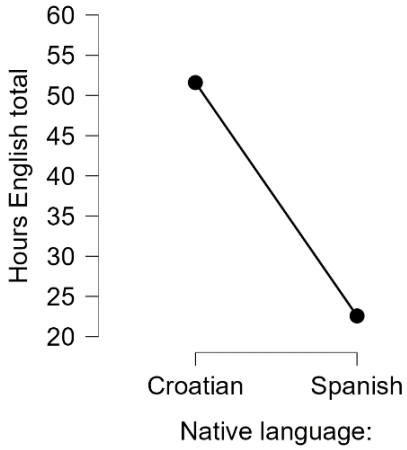


**Figure 5.** Descriprives plots: Hours English total and Native language

## 5. Conclusion

Some of the research hypotheses were confirmed and some were not. The results of the conducted inferential statictics tests supported the expectation that the proficiency level and the number of hours that the participants spend reading, listening and using the English language affect the number of English compound words they recognized. Although the research did not investigate compound word processing in a strict psycholinguistic environment, the sample was relatively small and there was a lack of time limit in the instrument, the data is representative of the population. The extent to which compound words are used, and their features in Croatian and Spanish is different, but what this research did not show is how it influences the perception of English compound words of the two groups of native speakers. It did, however, show other, external effects on compound word knowledge.

In conclusion, the aim of this paper was to present the morphological theory in language acquisition and explore the theoretical background of compound words and their acquisition in a second language. It explained the main concepts concerning morphological processing in psycholinguistics and showcased the main morphological processes that occur in inflection and derivation, as well as how the knowledge different languages affects acquiring other languages that enter the mental lexicon. A typology of English compound words was displayed, along with information about the process of their acquisition. After a short comparison of compounds in English, Spanish and Croatian, the paper delved into the methodology and analysis of the results from the conducted research. The research gave an insight into the influence of interaction with English-related leisure time activities on Croatian and Spanish native speakers' knowledge of English compound words. Although on a smaller scale, the results showed some significant results and assessed answering research questions. Hopefully, this paper managed to present the main postulates in language acquisition and compound words theory.

**References:**

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*. Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Archibald, J., & Libben, G. (2019). Morphological Theory and Second Language Acquisition. *The Oxford handbook of morphological theory*. Oxford University Press, UK.

Baayen, H. (2014). Experimental and psycholinguistic approaches to studying derivation. *The Oxford handbook of derivation*. Oxford University Press, UK.

Bell, M. J., & Schäfer, M. (2016). Modelling semantic transparency. *Morphology*. Springer, New York City, US.

Bisetto, A., & Scalise, S. (2005). The classification of compounds. *The Oxford handbook of compounding*. Lingue e linguaggio, Bologna, Italy.

Blom, E. (2019). Morphological Theory and First Language Acquisition. *The Oxford handbook of morphological theory*. Oxford University Press, UK.

Booij, G. (2007). The grammar of words: An introduction to linguistic morphology (Second Edition). Oxford University Press, UK.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language learning.* University of Essex, UK.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). *The role of the frequency of constituents in compound words: Evidence from Basque and Spanish*. Springer, New York City, US.

Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading research quarterly*. Wiley, New Jersey, US.

Fabb, N. (2017). Compounding. *The handbook of morphology*. Wiley, New Jersey, US.

Friedline, B. (2011). *Challenges in the second language acquisition of derivational morphology: from theory to practice.* University of Pittsburgh, Pennsylvania.

Grainger, J., & Dijkstra, T. (1992). *Cognitive processing in bilinguals*. Kansas State University Manhattan, Kansas.

Grainger, J.¸, & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology Section A*. Sage Publishing, Newbury Park, California.

Grčević, M. (2015). Croatian. *Word-Formation: An International Handbook of the Languages of Europe.* De Gruyter, Berlin, Germany.

Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research.* Sage Publishing, Newbury Park, California.

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language.* Elsevier, Amsterdam, Netherlands.

Libben, G. (1994). *Language and cognitive processes*. Taylor & Francis, Oxfordshire, UK.

Libben, G. (2006). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics.* Cambridge University Press, UK.

Libben, G. (2012). *Aspects of multilingual aphasia*. Multilingual matters, Bristol, UK.

Libben, G., & Goral, M. (2015). How bilingualism shapes the mental lexicon. *The Cambridge handbook of bilingual processing.* Cambridge University Press, UK.

Libben, G., Derwing, B. L., & de Almeida, R. G. (1999). Ambiguous novel compounds and models of morphological parsing. *Brain and Language*. Elsevier, Amsterdam, Netherlands.

Libben, G., Gagné, C. L. & Dressler, W. U. (2006). The representation and processing of compound words. *Word knowledge and word usage.* De Gruyter, Berlin, Germany.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). *Evolution of universal grammar*. Science.

Perlmutter, D. (1988). *Theoretical morphology*. Brill, Netherlands.

Piera, C. (1995). On compounding in English and Spanish. *Evolution and revolution in linguistic theory*. Georgetown University Press, Washington, D. C.

Renner, V. & Fernández-Domínguez, J. (2011). *Coordinate compounding in English and Spanish*. De Gruyter, Berlin, Germany.

Sebastián-Gallés, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). LEXESP: *Léxico informatizado del español*. Universitat de Barcelona, Spain.

Silva, R., & Clahsen, H. (2008). *Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English*. Cambridge University Press, UK.

Štambuk, A. (1997). *Multi-word lexical units in English and Croatian terminology of electronics*. University of Split, Croatia.

Steinke, S., & Zhai, W. (2020). *Transparency and Headedness in Processing Compound Words: A Lexical Decision Study*. Linguistic Portfolios. St. Cloud State University, Minnesota.

Taft, M. (1981). *Journal of verbal learning and verbal behavior*. Elsevier, Amsterdam, Netherlands.

Taft, M., & Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of verbal learning and verbal behavior*. Elsevier, Amsterdam, Netherlands.

Ullman, M. (2004). *Contributions of memory circuits to language: the declarative/procedural model.* Elsevier, Amsterdam, Netherlands.

Ullman, M. (2005). *A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model*. Georgetown University Press, Washington, D. C.

Wei, L. (2002). *The bilingual mental lexicon and speech production process.* Montclair State University, New Jersey, US.

**Appendix**

Questions that concerned interaction with English in leisure time were the following: 'How many hours a week do you spend watching movies or TV shows in English - original audio version?', 'How many hours a week do you spend watching dubbed movies or TV shows - translated from English to your native language?', 'How many hours a week do you spend reading books, articles or texts in English?', 'How many hours a week do you spend reading or watching content in English on social media?' and 'How many hours a week do you spend listening to music in English?'.

Questions in the compound word recognition test were listed as follows:

**1.** When we got back to the car there was a soggy parking ticket stuck on the _____.

1) windprotector

2) windcover

3) windshield

4) windguard

**2.** Because their job includes a lot of paperwork, lawyers always carry a _____.

1) quickcase

2) briefcase

3) briefbag

4) suitcase

**3.** To find out if you have a fever, place a thermometer under your _____ or use a special ear thermometer.

1) armpit

2) armhole

3) armdent

4) armgap

**4.** I prefer to switch off the odd electric light and read by _____ in bed.

1) candleshine

2) candleglow

3) candlebright

4) candlelight

**5.** _____ is a form of shorthand in which alphabetic combinations are used to represent groups of sounds or short common words.

1) fastwriting

2) speedwriting

3) shortwriting

4) quickwriting

**6.** The average _____ of the Forest Elephant is 60 years.

1) lifeage

2) lifestretch

3) lifeduration

4) lifespan

**7.** They were having lunch when they heard the _____.

1) doorbell

2) housering

3) housebell

4) doorring

**8.** I prefer to use loose tea leaves instead of _____ for a more authentic flavor.

1) teabags

2) teapacks

3) teasacks

4) teapackages

**9.** Perhaps a plastic organizing bin or a decorative _____ would be more fitting.

1) flowebowl

2) flowercan

3) flowerpot

4) flowerbucket

**10.** Sally keeps her new bird in a small _____.

1) birdcase

2) birdcage

3) birdbox

4) birdbag

**11.** In American football, the player who receives the ball at the start of every play and tries to move it along the field is called a _____.

1) quarterback

2) quarterfront

3) sideback

4) midfront

**12.** He is such a _____ - he doesn't like to give or spend money.

1) cheapslide

2) tightslide

3) cheapskate

4) poorskate

**13.** Yesterday I saw a huge green _____ in the vineyard.

1) grassjumper

2) grasshopper

3) grassbouncer

4) grassleaper

**14.** Every month the flatmates pay their rent to the person who owns the flat – the _____.

1) landowner

2) landholder

3) landkeeper

4) landlord

**15.** A model of a person dressed in old clothes and put in a field of growing crops to frighten birds away is called a _____.

1) scarecrow

2) fearcrow

3) scarebird

4) fieldscarer

**16.** Certainly not a _____, you're ready to dance the night away.

1) wallflower

2) floortree

3) sidewaiter

4) partyduddy

**17.** I'm afraid you're late - the _____ for applications was 30 May.

1) deadlimit

2) deadline

3) deadboundary

4) deadborder

**18.** Leaving the party so soon, he is such a _____.

1) nojoy

2) killjoy

3) zerojoy

4) missjoy

**19.** Are we going by your car?

- That's assuming my _____ is fit for the road.

1) bouncetrap

2) bumptrap

3) punchtrap

4) rattletrap

**20.** Don't do it like that, you _____!

1) numbskull

2) dumbskull

3) dullskull

4) slowskull

The SAT evidence-based sample test involved questions concerning the texts such as: 'What does Passage 1 suggest about the US government's provisions for the institution of slavery, as framed in the Constitution?' and 'Which choice provides the best evidence for the answer to the previous question?' for the first text and lexical decision multiple-choice questions for the second text such as: Which choice best fits with the tone of the rest of the passage: (there was a numbered reference in the text and the word in question was bold)

1) NO CHANGE
2) super-rigorous
3) spot-on
4) intense.

**Summary**

MORPHOLOGY IN SECOND LANGUAGE ACQUISITION: THE ACQUISITION OF ENGLISH COMPOUNDS

This paper deals with the morphological theory in second language acquisition with a focus on compound words. It delves into how language is processed inside the mind and explains the key systems of word storage and computation. The paper begins with an insight into morphological processing from a psychological point of view and discusses how different languages interfere in the speaker's mind. This part is followed by an overview of English compounds, their acquisition and the main aspects of compound words in Croatian and Spanish. Finally, it showcases the research conducted among speakers of English as a second language which investigates how well they know transparent and opaque English compound words and how their native language, proficiency level and leisure time activities affect it.

**Key words**: morphology, language acquisition, psycholinguistics, word processing, compound words

**Sažetak**

MORFOLOGIJA U USVAJANJU STRANOG JEZIKA: USVAJANJE ENGLESKIH SLOŽENICA

Ovaj rad se bavi teorijom morfologije u usvajanju stranog jezika s naglaskom na složenice. Ulazi u procesiranje jezika unutar uma te objašnjava osnovne sustave za pohranu i produkciju riječi. Ovaj rad započinje uvidom u morfološko procesiranje iz psiholingvističke perspektive i raspravlja o tome kako različiti jezici posreduju u govornikovom umu. Nakon toga, predstavlja se pregled složenica u Engleskom jeziku, njihovo usvajanje i glavni aspekti složenica u hrvatskom i španjolskom. Konačno, prikazuje istraživanje provedeno među govornicima engleskog kao stranog jezika koje ispituje koliko dobro poznaju transparentne i netransparentne engleske složenice te kako materinji jezik, znanje jezika i aktivnosti u slobodno vrijeme utječu na to.

**Ključne riječi**: morfologija, usvajanje jezika, psiholingvistika, procesiranje riječi, složenice